ACCSCIENCE PUBLISHING

Artificial Intelligence in Health

ORIGINAL RESEARCH ARTICLE

Enhancing COVID-19 severity assessment with artificial intelligence-based bone suppression technique in chest radiography

Asumi Yamazaki[1], Masashi Seki[2], and Takayuki Ishida[1]*

[1]Division of Health Sciences, Osaka University Graduate School of Medicine, Suita, Osaka, Japan

[2]Department of Radiology, Kitasato University Hospital, Sagamihara, Kanagawa, Japan

Abstract

Chest radiography (CXR) is widely used for initial respiratory assessment, but its lesion detection capability is typically inferior to that of computed tomography. Several studies have reported that artificial intelligence (AI)-based bone suppression techniques can enhance the accuracy of lesion detection and disease classification. Previously, we developed an AI-based bone suppression system based on dual-energy subtraction principles. However, the subtraction process limited its versatility and introduced significant artifacts. To overcome these challenges, we improved the system to generate bone-suppressed images directly, eliminating the need for subtraction. This study demonstrates the utility of the updated bone suppression system as a pre-processing tool for regression analyses in assessing coronavirus disease 2019 severity. Four regression models – DenseNet, ResNet18, ResNet50, and RegNetY-120 – were employed to predict the severity based on scores annotated by radiologists. Except for DenseNet, all models showed statistically significant improvements in Pearson correlation coefficients (PCCs) when using bone-suppressed images generated by the updated model. The highest PCC, 0.895, was achieved by the ResNet18 model. The direct image generation process improved the clinical practicality of the bone suppression system while reducing artifacts. Furthermore, the significant improvement in linearity suggests that AI-driven bone suppression enhances the visibility of abnormalities and improves the accuracy for pulmonary condition assessments. These advancements could expand the application of bone suppression techniques in various regression analyses, including disease severity, progression, and recurrence risk. Nonetheless, further validation using larger and more diverse datasets, as well as a broader range of prediction models, is necessary.

Keywords: Bone suppression; Artificial intelligence; COVID-19; Regression model; Chest radiography

***Corresponding author:**
Takayuki Ishida
(tishida@sahs.med.osaka-u.ac.jp)

Citation: Yamazaki A, Seki M, Ishida T. Enhancing COVID-19 severity assessment with artificial intelligence-based bone suppression technique in chest radiography. *Artif Intell Health*. 2025;2(3):95-106.
doi: 10.36922/aih.5608

Received: October 28, 2024

1st revised: January 10, 2025

2nd revised: February 11, 2025

Accepted: February 21, 2025

Published online: March 5, 2025

1. Introduction

Coronavirus disease 2019 (COVID-19) emerged in late 2019,[1] rapidly developing into a global pandemic that overwhelmed healthcare systems due to its high contagiousness, unprecedented morbidity, and mortality rates.[2,3] Since then, with the spread of various mutant strains, over seven million deaths have been reported worldwide.[4] While severe cases of COVID-19 can lead to death, many patients remain asymptomatic or experience

mild symptoms.[5] Mild cases are often self-limiting and do not require hospitalization, whereas severe cases necessitate admission to the intensive care unit; mechanical ventilation, including extracorporeal membrane oxygenation; and treatment with anti-inflammatory agents.[6] An early and accurate diagnosis, along with a reliable severity assessment, is critical for effective patient management and preventing the overburdening of healthcare facilities.

Chest radiography (CXR) is widely used for initial evaluations in medical emergencies due to its low cost, rapid examination, and widespread availability.[7] While CXR is useful for detecting COVID-19, particularly during pandemic situations, its sensitivity is lower than that of chest computed tomography (CT).[8,9] Typical imaging patterns observed in COVID-19 patients include bilateral, peripheral, and basal-predominant ground-glass opacities without pleural effusion.[8,10,11] Although CT is generally considered more reliable for identifying such subtle contrast lesions,[9] the potential of image pre-processing techniques to enhance COVID-19 detection using CXR has been extensively explored.[12-14] For instance, Sharifrazi *et al.*[12] demonstrated that Sobel filter pre-processing improves the performance of convolutional neural network (CNN) models for detecting COVID-19 using CXR. Similarly, Arias-Garzón *et al.*[13] reported that pre-processing with segmentation techniques leads to a more accurate and reliable COVID-19 classification. Avolio *et al.*[14] further suggested that pre-processing methods could benefit not only CNN-based approaches but also other machine learning techniques, such as multiple-instance learning.

Furthermore, Takaki *et al.*[15] demonstrated that bone suppression processing through temporal subtraction significantly increased radiologists' sensitivity and reduced their false-positive rates in detecting pulmonary lesions on CXR. In contrast, van der Heyden[16] suggested that dual-energy subtraction (DES) technology enhances the diagnostic accuracy of CXR by eliminating bone structures and improving soft-tissue visualization. Building on these conventional bone suppression methods, advancements in artificial intelligence (AI) have introduced innovative alternatives for enhancing bone suppression.[17-30] Unlike conventional DES systems, which rely on dual X-ray exposure and specialized detectors equipped with copper plates, AI-based DES (AI-DES) approaches improve accessibility while offering additional advantages, such as reduced motion artifacts and improved noise characteristics. A representative study employed a generative adversarial network (GAN)-based model to suppress bone structures in CXR without the need for labeled data, utilizing digitally reconstructed radiographs from CT data.[18] Recent advancements have further extended these applications to pediatric imaging.[19,20] For example, Xie *et al.*[20] utilized bone edge detection to generate both bone-enhanced and bone-suppressed images, demonstrating the potential for improved pneumonia diagnosis in children.

We also previously developed an AI-DES system that successfully generated bone-suppressed images; however, its clinical applicability was limited by the need for raw high- and low-energy images, which are often unavailable, and by labor-intensive weighted subtraction processing.[22] To overcome these limitations, we updated the system to generate bone-suppressed images directly from routine chest radiographs in this study. This improvement enhances the system's versatility and clinical utility by streamlining the entire image-processing workflow.

Several studies have demonstrated the advantages of AI-based bone suppression techniques in classification tasks related to various lung diseases, including COVID-19.[23-29] In particular, Rani *et al.*[23] proposed a model that preserves spatial features, suggesting that effectively pre-processed radiographs could enhance diagnostic performance. The authors further reported that integrating AI-based bone suppression with pre-processing techniques, such as lung segmentation and augmentation, significantly improves the classification accuracy of pneumonia, including COVID-19.[24] Lam *et al.*[25] reported that bone-suppressed images, generated using a CNN-based model proposed by Rajaraman *et al.*,[26] significantly increased the area under the receiver operating characteristic curve for COVID-19 classification tasks, compared to standard radiographs using a modified VGG16 model. Similarly, Xu *et al.*[30] developed a CNN-based rib removal model, SADXNet, which showed superior performance in lung nodule detection, lung anomaly classification, and localization tasks. These findings highlight the importance of bone suppression in enhancing the accuracy of lung disease classification.

Given these promising results, we hypothesize that bone suppression techniques may also be effective in regression tasks, such as assessing disease severity, predicting the risk of progression, and estimating patient prognosis. Regression models, which predict continuous values by identifying subtle variations, can be more complex and challenging to construct than classification models, particularly when dealing with high-dimensional or imbalanced data.[31] This study explores the utility of AI-based bone suppression in COVID-19 severity assessment using CXR, comparing the performance of regression models with and without bone suppression to validate its effectiveness.

Recent studies have proposed AI-based prediction models for COVID-19 severity assessment.[32-34] Cohen *et al.*[32,33] used DenseNet-based regression models to evaluate COVID-19 severity based on the extent of lung

involvement and the degree of opacity, achieving Pearson correlation coefficients (PCCs) of 0.80 and 0.78 for these tasks, respectively. Signoroni *et al.*[34] introduced BS-Net, an end-to-end architecture, to segment, align, and quantify lung compromise based on the Brixia score.[35,36] The performance of BS-Net was evaluated not only for classification tasks but also for regression tasks using linear regression of the Brixia score, with the highest PCC reaching 0.85. These studies underscore the potential of AI approaches in evaluating COVID-19 severity.

Moreover, transparency, explainability, and interpretability are critical components of AI, especially in medical applications.[37] Understanding why and how a model derives a particular decision is essential for ensuring clinical accountability and building confidence. Gradient-weighted class activation mapping (Grad-CAM) is a widely used technique to visualize decision-making processes by highlighting image regions that contribute to the model's output.[38] By generating heatmaps based on gradients from the final convolutional layer, Grad-CAM offers explainable insights to support clinical decision-making.[39,40] Talaat *et al.*[40] integrated Grad-CAM into a breast cancer classification model, providing radiologists with valuable insights into the model's decision-making process and fostering trust in the AI system. In this study, Grad-CAM is used to validate the explainability and interpretability of COVID-19 severity prediction models.

Our work integrates AI-based bone suppression pre-processing into regression models to assess COVID-19 severity using CXR. The primary aim is to expand the applications of AI-based bone suppression techniques, verifying their utility in severity assessment. By improving the accuracy of severity predictions, this approach could enhance patient monitoring and optimize healthcare resource allocation, particularly in resource-limited settings. Our findings may also validate the applications of AI-based bone suppression in regression tasks for chest image diagnosis. Moreover, this study seeks to bridge the gap between the present limitations of CXR and the superior sensitivity of CT, ultimately contributing to more efficient and scalable diagnostic tools for COVID-19 and other pulmonary diseases.

2. Data and methods

In this section, we explain the development of the bone suppression model, followed by the method for assessing COVID-19 severity.

2.1. Bone suppression model

2.1.1. Data collection

We collected chest radiographs from 600 patients using a dual-shot DES system (Discovery XR656, GE Healthcare, Chicago, IL, USA) at Kitasato University Hospital (Sagamihara City, Japan), to develop a bone suppression model. Most of these patients had pulmonary inflammatory diseases or pulmonary mass lesions. The detector specifications are detailed in our previous work.[22] Radiography was performed with tube voltages of 130 kV for high-energy images and 60 kV for low-energy images. The system produces bone-suppressed and bone-enhanced images, along with standard chest radiographs for presentation, all with a resolution of 3524 × 4288 pixels and 13-bit contrast, from the raw data of the high- and low-energy images. For training, we utilized 480 pairs of standard and bone-suppressed radiographs, while 120 pairs were reserved for testing.

2.1.2. Data pre-processing

To prepare the dataset for model training, we first cropped the standard and bone-suppressed radiographs to extract regions of interest (ROIs) centered on the lung area. The lung regions were identified using a pre-trained U-Net[41] model, which segments chest radiographs into the lung, heart, other anatomical areas, and background, assigning pixel values of 255, 85, 170, and 0, respectively, in 8-bit contrast. The U-Net architecture employed consisted of five depths, incorporating an input layer, five encoder layers, five decoder layers, and an output layer.

For training the U-Net model, we utilized all 247 chest radiographs from the Japanese Society of Radiological Technology database, along with their corresponding segmented labels.[42] The U-Net model was trained for up to 100 epochs using the RMSprop optimizer, with a learning rate of 0.0001, a weight decay of 1×10^{-8}, and a momentum of 0.9.

After training, the U-Net model was applied to identify the lung regions in the standard radiographs collected at Kitasato University Hospital, which had been converted to 8-bit contrast in advance. These identified lung regions were then cropped from both the standard and bone-suppressed radiographs. Finally, the cropped images were resized to 1024 × 1024 pixels to standardize the input size for the subsequent training of the bone suppression model.

2.1.3. Bone suppression network architecture and training settings

We employed the pix2pix[43,44] network to generate virtually bone-suppressed images from the standard chest radiographs. Figure 1 illustrates a flowchart of the bone suppression and pre-processing steps. The network architecture follows the design proposed by Isora *et al.*,[43] as described in our previous work,[22] with modifications made to the resolution of the generator and discriminator to handle 1024 × 1024 resolution images.

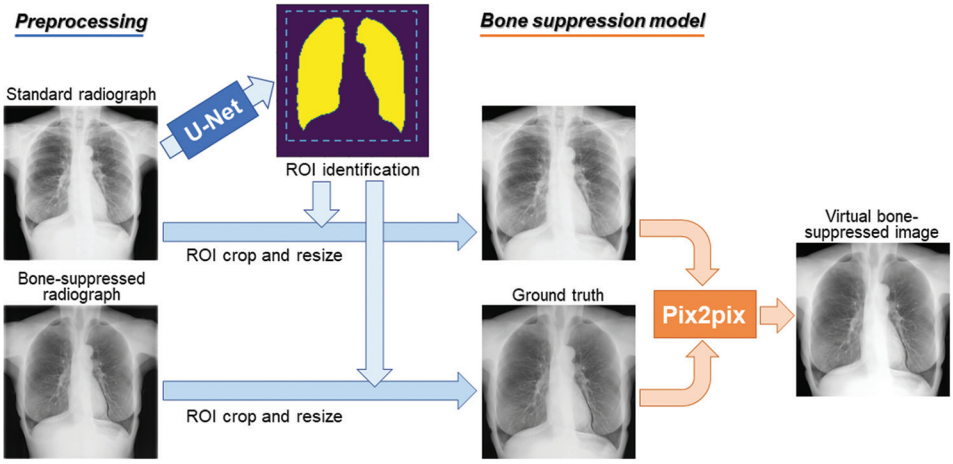

Figure 1. Flowchart of bone suppression model training
Abbreviation: ROI: Region of interest.

The input datasets in 8-bit contrast were normalized by converting pixel values to floating-point values ranging from 0 to 1. This normalization allowed the pix2pix network to process floating-point data, effectively reducing quantization errors. After processing, the network output was rescaled by multiplying the floating-point values by 255, converting them back to 8-bit images.

The network was trained for up to 4000 epochs, with a batch size of 2, using an NVIDIA TITAN RTX on an Ubuntu 20.04.6 LTS operating system. We applied the Adam optimizer with momentum parameters of $\beta 1 = 0.5$ and $\beta 2 = 0.999$. The learning rates were dynamically adjusted throughout the training: The generator started at 0.002 and decreased by 0.002 for each epoch, while the discriminator started at 0.02 and decreased by 0.02 per epoch. The implementation was carried out using Python 3.9.18 and PyTorch 1.12.0.

2.1.4. Performance evaluation

The performance of the bone suppression model was evaluated using the test dataset by measuring the similarity between the generated virtual bone-suppressed images and the ground truth images. Image similarity was assessed using two metrics: The peak signal-to-noise ratio (PSNR) and the structural similarity index (SSIM).[45,46]

The PSNR measures image similarity based on the ratio of noise to the maximum pixel values, calculated as follows:

$$PSNR = 20 \bullet \log_{10}\left(\frac{P_{max}}{MSE}\right), \quad \text{(I)}$$

Where mean squared error *(MSE)* is the mean square error between two images, and P_{max} is the maximum value, which is 255 in this study.

The SSIM measures image similarity between two images, x and y, as follows:

$$SSIM = \frac{\left(2\mu_x\mu_y + C_1\right)\left(2\sigma_{xy} + C_2\right)}{\left(\mu_x^2 + \mu_y^2 + C_1\right)\left(\sigma_x^2 + \sigma_y^2 + C_2\right)}, \quad \text{(II)}$$

Where μ_x and μ_y are the means; σ_x^2 and σ_y^2 are the variances; and σ_{xy} is the covariance for x and y in the ROIs, which consist of 5 × 5 pixels in this study. C_1 and C_2 are defined as $C_1 = (0.01L)^2$ and $C_2 = (0.03L)^2$, where L is 255 in this study.

2.2. COVID-19 severity assessment model

2.2.1. Data collection

We selected 192 chest radiographs from 136 COVID-19 patients in a publicly available image database provided by Cohen *et al.*[33] Although the database contains more than 700 images collected in several medical centers across 26 different countries, only 192 images were annotated using the Brixia score[33,34] by two expert radiologists: A board-certified specialist with 22 years of experience and a trainee with 2 years of experience. The scoring system was initially introduced in a radiology department in Italy during the pandemic, and it was later validated for risk stratification in a large population.[36] This score evaluates pneumonia severity by dividing the lungs into six zones and assigning an integer score ranging from 0 to 3 to each zone. Specifically, a score of 0 denotes no lung abnormalities, 1 indicates the presence of interstitial infiltrates, 2 reflects a combination of interstitial (dominant) and alveolar infiltrates, and 3 signifies interstitial and alveolar (dominant) infiltrates.

The total Brixia score, which ranges from 0 to 18, was calculated by summing the scores of all six zones. To

derive the final labels for the regression models described in subsection 2.2.3., we averaged the total scores from both radiologists and normalized this value to a floating-point number between 0 and 1 by dividing by 18. These normalized scores were used as the labels for the regression models. The mean and standard deviation (SD) of the scores across 192 images were 0.380 and 0.260, respectively. Figure 2 provides an overview of the severity assessment process.

2.2.2. Data pre-processing

First, the radiographs were cropped and resized to center on the lung area, following the process described in subsection 2.1.2. Next, the images were transformed into bone-suppressed images using the AI-based bone suppression model developed in subsection 2.1. Both the standard radiographs and the bone-suppressed images were independently transformed to a resolution of 512 × 512 pixels with 8-bit contrast. Subsequently, each type of image was then randomly split into training and test data in an approximately 80:20 ratio, ensuring that all images from the same patient were grouped together in the same split. Five-fold cross-validation was applied to each dataset separately, and, to enhance robustness, this process was repeated 3 times using different random seeds.

2.2.3. Regression models and training settings

We employed several CNN architectures from different generations – DenseNet,[47] ResNet18, ResNet50,[48] and RegNetY-120[49] – all pre-trained on ImageNet.[50] To adapt these models for the regression task, we modified their final fully connected layers to have a single output that predicts a continuous value, corresponding to the normalized Brixia score. These models were trained using the MSE loss function for up to 25 epochs in each cross-validation fold. The training was conducted on an NVIDIA GeForce RTX 4070 with a Windows 11 operating system, utilizing Python 3.8.18 and PyTorch 2.2.1.

Based on our initial experiments, which indicated that the Stochastic Gradient Descent (SGD) optimizer consistently outperformed the Adam optimizer, we adopted SGD with a learning rate of 0.001 and a momentum of 0.9 for all models. In addition, a learning rate scheduler (StepLR) was applied to reduce the learning rate by a factor of 0.1 every 5 epochs.

2.3. Performance evaluation

2.3.1. Relationship between truths and predictions

We compared the performance of the severity assessment models between the standard chest radiograph dataset and the bone-suppressed image dataset by computing the mean ± SD of the mean absolute errors (MAEs) and PCCs across all folds and random seeds. The PCC quantifies the linear relationship between two variables,[51] as expressed by the following:

$$r = \frac{\sum_{i=0}^{n-1}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=0}^{n-1}(x_i - \bar{x})^2}\sqrt{\sum_{i=0}^{n-1}(y_i - \bar{y})^2}}, \quad \text{(III)}$$

Where r is the PCC; x_i and y_i denote the individual sample points; and $\bar{x}$ and $\bar{y}$ are the means of x_i and y_i, respectively.

These metrics were calculated using the "mean_absolute_error" function from the Python "sklearn.metrics" library and the "pearsonr" function from the Python "scipy.stats" module.

Statistical significance tests were conducted using a two-tailed Student's t-test to compare the average MAEs

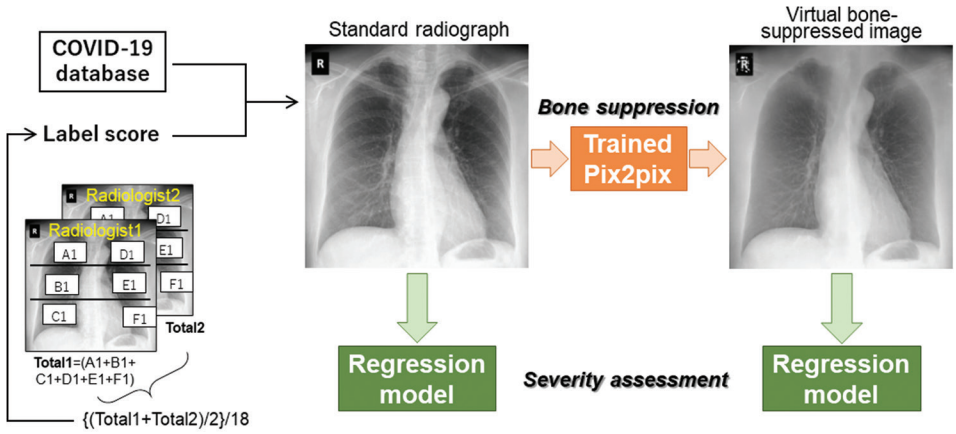

Figure 2. Flowchart of coronavirus disease 2019 severity assessment. The Brixia scoring system assigns an integer value from 0 to 3 to each of the six lung zones (A1 to F1). The total scores from two radiologists were averaged and then normalized to generate the final label scores.

and PCCs between the standard chest radiograph dataset and the bone-suppressed image dataset. The statistical tests and calculation of *P*-values were performed using the Python "scipy.stats" module.

2.3.2. Explainability of the severity assessment models

To validate the explainability and interpretability of the severity assessment models, we applied Grad-CAM[38] to generate heatmaps that exhibit the gradients in the final convolutional layer for the corresponding datasets tested in subsection 2.3.1. We used the "visualize_cam" function from the "gradcam.utils" module to generate these heatmaps, highlighting the regions that are most influential in predicting the severity.

3. Results

3.1. Generated bone-suppressed images

Figure 3 presents the bone-suppressed images generated by our updated bone suppression model, compared with the corresponding ground truth images for three cases from the test dataset collected at Kitasato University Hospital. The generated images closely resemble the ground truth, exhibiting a high degree of image similarity, with an average PSNR of 40.4 dB and an SSIM of 0.962 across the entire test dataset. Effective bone suppression was particularly achieved in the ribs and vertebral bones while preserving pneumonia and mass lesions.

In our previous AI-DES model, insufficient bone suppression was an avoidable issue due to enhanced quantization errors in the subtraction process.[22] In contrast, the updated model shows a significant improvement in bone suppression by directly generating bone-suppressed images, eliminating the need for the subtraction process. Furthermore, as shown in the enlarged images in Figure 3, the ground truth image of the third case exhibits motion artifacts, whereas the generated image displays a remarkable reduction in these artifacts. These findings highlight the model's ability to enhance image quality, surpassing that of the ground truth and our previous model.

Figure 4 showcases four examples of standard chest radiographs from the COVID-19 database, accompanied by the corresponding bone-suppressed images generated by our bone suppression model, and their severity score labels based on Brixia scores. This demonstrates the robust effectiveness of our bone suppression model, even when applied to an external dataset with diverse lung conditions.

3.2. Performance in COVID-19 severity assessment

Table 1 compares the performance of each trained regression model on the standard chest radiograph dataset versus the bone-suppressed image dataset, showing the averages and SDs of the MAEs and PCCs for the test data across all folds and random seeds. The table also includes the results of statistical significance tests. For cases where statistically significant differences were observed ($P < 0.05$), the better-performing averages are highlighted in bold, along with the corresponding *P*-values.

The ResNet18, ResNet50, and RegNetY-120 models demonstrated statistically significant improvements in the PCCs for the bone-suppressed image dataset compared to the standard chest radiograph dataset. In addition, the ResNet18 and RegNetY-120 models exhibited statistically significant lower MAEs, indicating superior predictive performance. In contrast, the DenseNet model showed similar performance on both datasets, with no statistically significant differences in either the MAEs or PCCs.

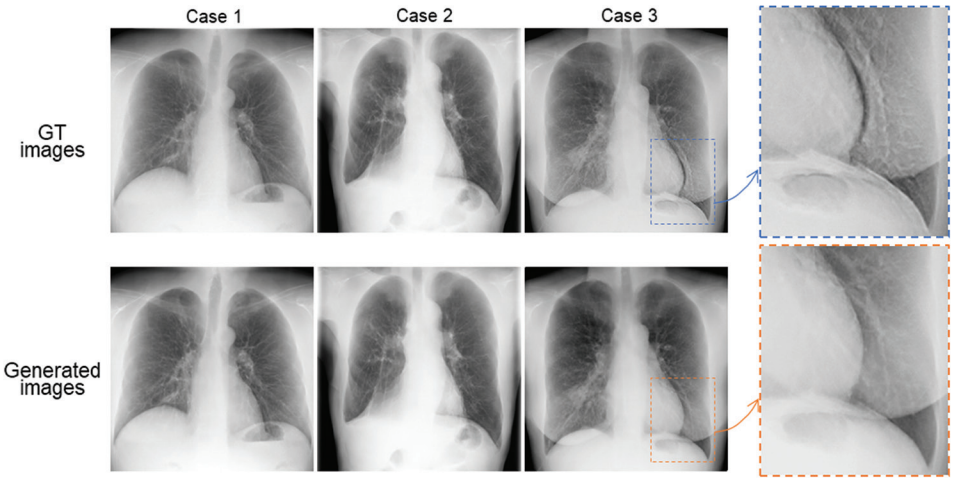

Figure 3. Comparison of virtually generated bone-suppressed images and the ground truth images. The third case also presents the enlarged images of the lower left lung field.

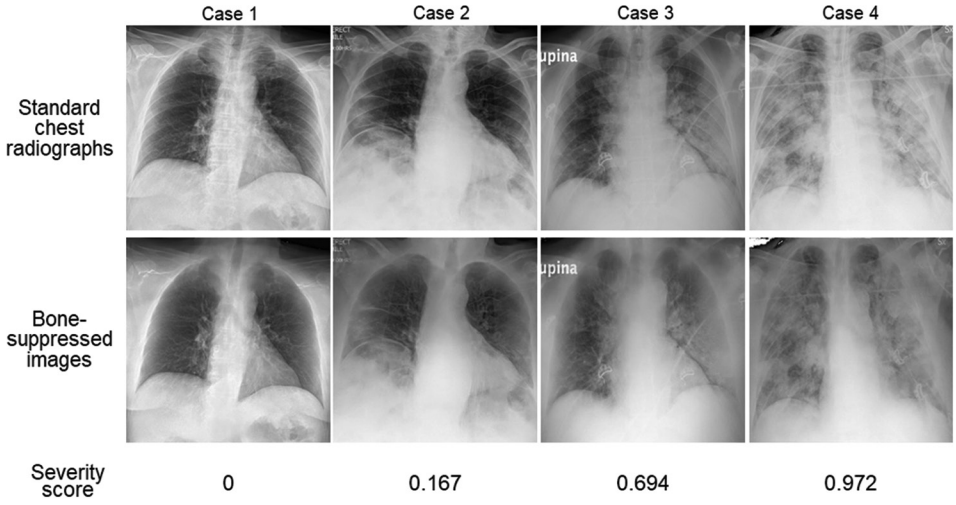

Figure 4. Examples of standard chest radiographs from a publicly available coronavirus disease 2019 database, along with the corresponding bone-suppressed images produced by our artificial intelligence model. The severity scores based on the Brixia score are displayed at the bottom of each image pair.

Table 1. Performance metrics of each model for standard chest radiographs versus bone-suppressed images in COVID-19 severity assessment

Model	MAE			PCC		
	Standard radiograph	BS image	Statistical significance	Standard radiograph	BS image	Statistical significance
DenseNet	0.0823±0.0286	0.0768±0.0451	NS	0.864±0.0953	0.873±0.164	NS
ResNet18	0.112±0.0551	**0.0722**±0.0435	0.00225*	0.797±0.216	**0.895**±0.150	0.0175*
ResNet50	0.0843±0.0261	0.0685±0.0178	NS	0.858±0.130	**0.882**±0.105	0.0230*
RegNetY-120	0.112±0.0379	**0.0993**±0.0319	0.0448*	0.811±0.143	**0.868**±0.0984	0.0181*

Note: Means±standard deviations across all folds are presented for the MAEs and PCCs. The statistical significance column lists *P* values for statistically significant cases (P<0.05*). Value in boldface indicated the better-performing average.
Abbreviations: BS: Bone-suppressed; MAE: Mean absolute error; NS: not significant; PCC: Pearson correlation coefficient, COVID-19: Coronavirus disease 2019.

Figures 5 and 6 illustrate examples of standard chest radiographs, bone-suppressed images, their corresponding Grad-CAM-generated heatmaps, and the scores predicted by the ResNet50 model. In the heatmaps, the color spectrum represents activation levels, with red indicating the highest activation areas, followed by yellow, blue, and transparency as activation decreases. In most cases in the test dataset, the high-activation areas were relatively more focused on the inner lung regions in the bone-suppressed images than in the standard radiographs, as shown in Figures 5 and 6. Furthermore, in the two cases in Figure 5, the predicted scores from the bone-suppressed images were closer to the true score labels than those from the standard radiographs. For example, in case 1, labeled with a true score of 0.556, the severity score predicted from the bone-suppressed image was 0.537, while the score from the standard radiograph was 0.342. In contrast, the two cases in Figure 6 illustrate instances where the predicted scores from the bone-suppressed images are deviated further from the true labels than those from the standard radiographs, despite the heatmaps consistently indicating high-activation areas in the lung regions. In particular, in case 2, which has a true score of 0.472, the heatmap for the bone-suppressed image indicates the highest activation in the right lung area; however, the predicted score of 0.401 was further from the true label than the score of 0.493 predicted from the standard radiograph.

4. Discussion

In this study, we developed an AI-based bone suppression model for CXR and applied it to a publicly available COVID-19 image database. The pix2pix model demonstrated a high degree of image similarity to the ground truth images, achieving PSNR and SSIM metrics comparable to those reported by existing bone suppression models for chest radiographs.[17,23,26-27] As a result, our present model effectively removes bone structures while enhancing the visibility of lung tumors and inflammation

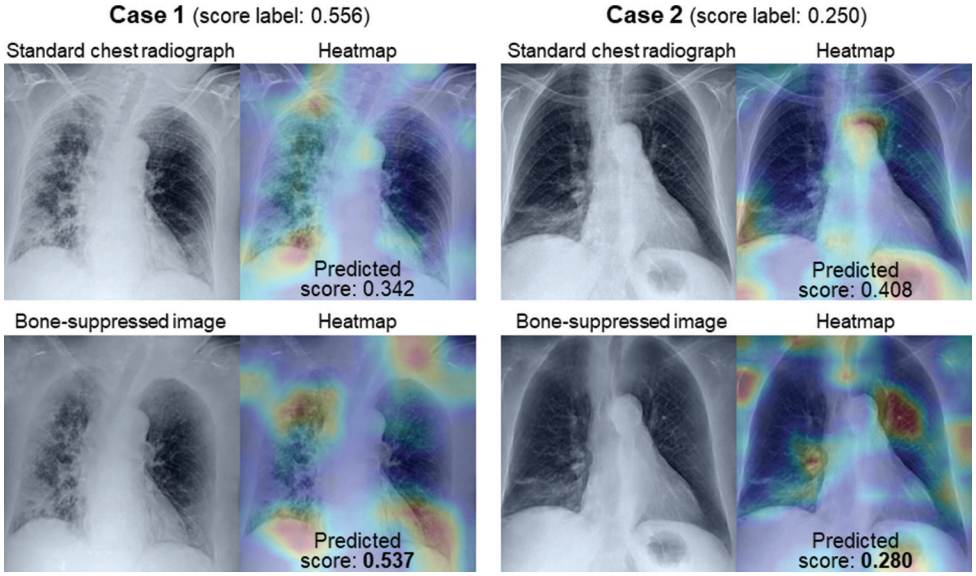

Figure 5. Standard chest radiographs, bone-suppressed images, their corresponding heatmaps, and the scores predicted by the ResNet50 model for two test cases, where the predicted scores from the bone-suppressed images are closer to the true score labels than those from the standard radiographs.

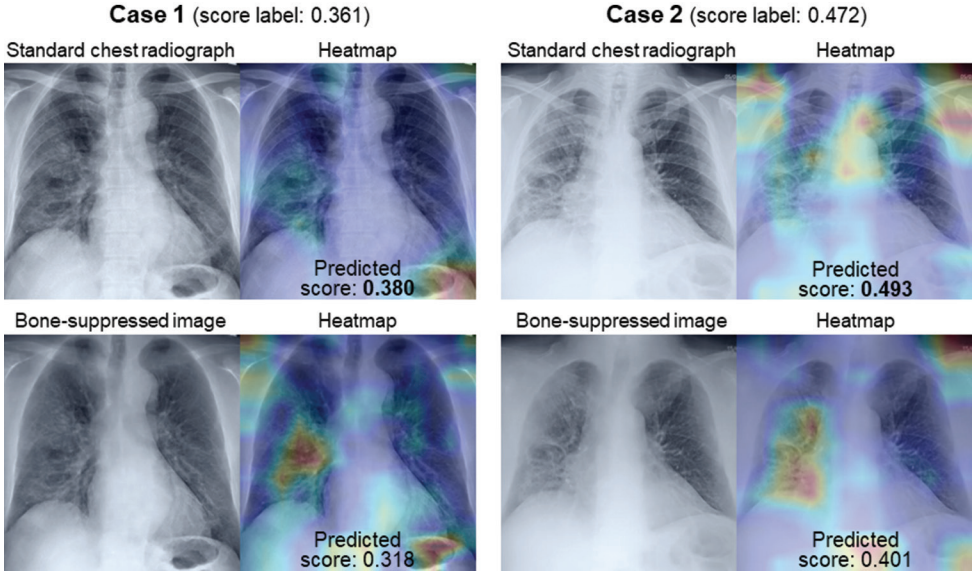

Figure 6. Standard chest radiographs, bone-suppressed images, their corresponding heatmaps, and the scores predicted by the ResNet50 model for two test cases, where the predicted scores derived from the standard radiographs are closer to the true score labels than those from the bone-suppressed images.

and reducing motion artifacts, addressing the limitations of traditional dual-shot DES systems.[52,53] Further improvements to the model architecture and parameter optimization could enhance its performance. For instance, Rani *et al.*[23] combined the pix2pix discriminator with Wasserstein GAN with gradient penalty,[54] achieving higher similarity with a PSNR of 43.588 and an SSIM of 0.989.

The enhanced image quality and practicality of our updated system, which eliminates the need for labor-intensive subtraction processing, effectively address the limitations of our previous system.[22] Moreover, this improved model has the potential to significantly enhance the diagnostic capabilities of CXR while maintaining the cost-effectiveness and time-efficiency benefits over CT scans. Its clinical applicability is further supported by its ability to generate high-quality images when applied to an external dataset, suggesting robustness across diverse scenarios, including different races, clinical conditions, and imaging systems. However, the limited sample size of the COVID-19 images used in this study underscores the

need for further validation, as AI model performance is often influenced by biases in the training data. Specifically, the small dataset size may have led to imbalances in disease severity, patient age distribution, and gender ratio. In fact, the mean label score of 0.380 and SD of 0.260 suggest that most of the images are associated with lower severity scores, potentially limiting the model's performance and generalizability. Consequently, the effectiveness of bone suppression for severity assessment may not be applicable to all patient populations or clinical settings. Furthermore, artificially generated images can exhibit undesirable artifacts or implausible shadows, even when produced by cutting-edge models.[55] Therefore, the quality of the bone-suppressed images should be carefully inspected by experts, such as radiologists and respiratory medicine physicians, to identify potential artifacts and ensure that the visibility of lesions is preserved.

We applied the updated bone suppression model as a pre-processing step in COVID-19 severity assessment using regression models. By utilizing the bone-suppressed images, we observed statistically significant improvements in linear correlations between predicted scores and true labels across various regression models, including classical models such as ResNet18 and the newer model RegNetY-120. Notably, the highest PCC of 0.895 exceeded the performance reported in related studies.[32,34] However, the DenseNet model showed no significant differences in performance with and without bone suppression. We acknowledge that, again, the small dataset size remains a significant limitation, potentially affecting the generalizability of this study. Furthermore, the variability in performance across different regression models warrants further investigation. In future work, we plan to confirm the effectiveness and robustness of bone suppression techniques using larger and more diverse datasets and a broader range of prediction models.

The clinical implications of this study are promising. The proposed bone suppression system has the potential to enhance the diagnostic performance of CXR by improving the visibility of abnormalities and enabling more accurate disease assessment. Beyond its established effectiveness in classification tasks, bone suppression techniques could facilitate regression-based evaluations of disease severity, progression, and recurrence risk. Integrating this system into diagnostic image viewing software would allow radiologists and physicians to access bone-suppressed images as additional clinical information. Moreover, as demonstrated in our assessment of COVID-19 severity, these techniques could serve as a pre-processing tool for computer-aided diagnosis systems. This integration may help reduce the discrepancy in diagnostic accuracy between CXR and CT.

In addition, we employed Grad-CAM to visualize the rationale for the model's decisions. We confirmed that the bone suppression techniques effectively directed the model's focus to the inner lung field. The effect could enhance the explainability and interpretability of the models, thereby increasing confidence in their predictions. However, even when the activated areas of the models were focused within the lung fields, the predicted scores did not consistently align with the true scores. Consequently, improving prediction accuracy while fostering explainability remains a challenge for achieving broader acceptance as a reliable diagnostic support tool.

In summary, this study improved the clinical applicability of the AI-based bone suppression system by eliminating the need for subtraction processing. The updated system enhanced the visibility of lung abnormalities, leading to more accurate predictions of pneumonia severity, as demonstrated by statistically significant improvements in the linear correlation between predicted severity scores and actual labels. These findings underscore the utility of AI-driven bone suppression in CXR, particularly for regression tasks related to disease severity, progression, and recurrence risk. Furthermore, the bone suppression techniques guided the prediction model to concentrate on the inner lung fields, suggesting potential improvements in the reliability of clinical assessments. The application of bone suppression in assessing COVID-19 severity could optimize patient monitoring and healthcare resource allocation. In addition, this advancement has the potential to elevate the diagnostic accuracy of CXR, providing valuable tools to overcome existing limitations, such as inferior contrast resolution and the superimposition of anatomical structures compared to CT.

5. Conclusion

This study successfully developed and validated an AI-based bone suppression model for CXR, which effectively removes bone structures while highlighting lung abnormalities. The model not only improves image quality but also streamlines the entire image processing workflow, increasing its clinical practicality. We applied the bone suppression model as a pre-processing step to facilitate more accurate predictions of COVID-19 severity. The findings demonstrate the potential of bone suppression techniques in assessing various pulmonary conditions, particularly through regression analyses. Future research should focus on validating bone suppression techniques with larger and more diverse datasets, as well as exploring a range of prediction models. In addition, addressing potential biases in AI outputs and enhancing the model's explainability will be essential for ensuring its reliable integration into routine clinical practice.

Acknowledgments

None.

Funding

This work was supported by the Kondou Kinen Medical Foundation.

Conflict of interest

The authors declare they have no competing interests.

Author contributions

Conceptualization: Asumi Yamazaki, Takayuki Ishida
Formal analysis: Asumi Yamazaki
Investigation: Asumi Yamazaki, Masashi Seki
Methodology: Asumi Yamazaki
Writing – original draft: Asumi Yamazaki
Writing – review & editing: Masashi Seki, Takayuki Ishida

Ethics approval and consent to participate

This study involved human subjects and was approved by the Research Ethics Review Committee of Kitasato University Hospital (approval number: C22-064) and Osaka University Graduate School of Medicine (approval number: 22061-6).

Consent for publication

A waiver of informed consent was granted by the Research Ethics Review Committee of Kitasato University Hospital and Osaka University Graduate School of Medicine due to the retrospective nature of this study. In accordance with Japan's Ethical Guidelines for Medical and Health Research Involving Human Subjects, patients were provided with the opportunity to "opt out."

Availability of data

The imaging data are not available due to institutional and ethical restrictions. The publicly available dataset can be accessed through the references cited in this article.

References

1. World Health Organization. *Coronavirus Disease* (*COVID-19*) *Pandemic*. Available from: https://www.who.int/emergencies/diseases/novel-coronavirus-2019 [Last accessed on 2024 Oct 15].
2. Beeraka NM, Sukocheva OA, Lukina E, Liu J, Fan R. Development of antibody resistance in emerging mutant strains of SARS CoV-2: Impediment for COVID-19 vaccines. *Rev Med Virol.* 2022;32(5):e2346.

   doi: 10.1002/rmv.2346
3. Ramirez FE, Sanchez A, Pirskanen A.T. Hydrothermotherapy in prevention and treatment of mild to moderate cases of COVID-19. *Med Hypotheses*. 2021;146:110363.

   doi: 10.1016/j.mehy.2020.110363
4. World Health Organization. *WHO Coronavirus* (*COVID-19*) *Dashboard*. Available from: https://data.who.int/dashboards/covid19/deaths?n=c [Last accessed on 2024 Oct 15].
5. Gao Z, Xu Y, Sun C, *et al*. A systematic review of asymptomatic infections with COVID-19. *J Microbiol Immunol Infect*. 2021;54(1):12-16.

   doi: 10.1016/j.jmii.2020.05.001
6. Kim MS, An MH, Kim WJ, Hwang TH. Comparative efficacy and safety of pharmacological interventions for the treatment of COVID-19: A systematic review and network meta-analysis. *PLoS Med*. 2020;17(12):e1003501.

   doi: 10.1371/journal.pmed.1003501
7. Long B, Carius BM, Chavez S, *et al*. Clinical update on COVID-19 for the emergency clinician: Presentation and evaluation. *Am J Emerg Med*. 2022;54:46-57.

   doi: 10.1016/j.ajem.2022.01.028
8. Wong HYF, Lam HYS, Fong AH, *et al*. Frequency and distribution of chest radiographic findings in patients positive for COVID-19. *Radiology*. 2020;296(2):E72-E78.

   doi: 10.1148/radiol.2020201160
9. Sverzellati N, Ryerson CJ, Milanese G, *et al*. Chest radiography or computed tomography for COVID-19 pneumonia? Comparative study in a simulated triage setting. *Eur Respir J*. 2021;58(3):2004188.

   doi: 10.1183/13993003.04188-2020
10. Tang JS, Lai JK, McCusker MW, Irving L, Pascoe DM, Heinze SB. Chest imaging findings in COVID-19-positive patients in an Australian tertiary hospital. *J Med Imaging Radiat Oncol*. 2022;66(6):755-760.

    doi: 10.1111/1754-9485.13339
11. Landini N, Colzani G, Ciet P, *et al*. Chest radiography findings of COVID-19 pneumonia: A specific pattern for a confident differential diagnosis. *Acta Radiol*. 2022;63(12):1619-1626.

    doi: 10.1177/02841851211055163
12. Sharifrazi D, Alizadehsani R, Roshanzamir M, *et al*. Fusion of convolution neural network, support vector machine and Sobel filter for accurate detection of COVID-19 patients using X-ray images. *Biomed Signal Process Control*. 2021;68:102622.

    doi: 10.1016/j.bspc.2021.102622
13. Arias-Garzón D, Alzate-Grisales JA, Orozco-Arias S, *et al*. COVID-19 detection in X-ray images using convolutional neural networks. *Mach Learn Appl*. 2021;6:100138.

    doi: 10.1016/j.mlwa.2021.100138

14. Avolio M, Fuduli A, Vocaturo E, Zumpano E. A comparative study of linear type multiple instance learning techniques for detecting COVID-19 by chest X-ray images. *Prog Artif Intell*. 2024.

doi: 10.1007/s13748-024-00332-1

15. Takaki T, Murakami S, Tani N, Aoki T. Evaluation of the clinical utility of temporal subtraction using bone suppression processing in digital chest radiography. *Heliyon*. 2023;9(1):e13004.

doi: 10.1016/j.heliyon.2023.e13004

16. Van der Heyden B. The potential application of dual-energy subtraction radiography for COVID-19 pneumonia imaging. *Br J Radiol*. 2021;94(1120):20201384.

doi: 10.1259/bjr.20201384

17. Matsubara N, Teramoto A, Saito K, Fujita H. Bone suppression for chest X-ray image using a convolutional neural filter. *Australas Phys Eng Sci Med*. 2019;43:97-108.

doi: 10.1007/s13246-019-00822-w

18. Han L, Lyu Y, Peng C, Zhou SK. GAN-based disentanglement learning for chest X-ray rib suppression. *Med Image Anal*. 2022;77:102369.

doi: 10.1016/j.media.2022.102369

19. Cho K, Seo J, Kyung S, Kim M, Hong GS, Kim N. Bone suppression on pediatric chest radiographs via a deep learning-based cascade model. *Comput Methods Programs Biomed*. 2022;215:106627.

doi: 10.1016/j.cmpb.2022.106627

20. Xie W, Gan M, Tan X, Li M, Yang W, Wang W. Efficient labeling for fine-tuning chest X-ray bone-suppression networks for pediatric patients. *Med Phys*. 2024;52:978-992.

doi: 10.1002/mp.17516

21. Ait Nasser A, Akhloufi MA. A review of recent advances in deep learning models for chest disease detection using radiography. *Diagnostics* (*Basel*). 2023;13(1):159.

doi: 10.3390/diagnostics13010159

22. Yamazaki A, Koshida A, Tanaka T, Seki M, Ishida T. Development of artificial intelligence-based dual-energy subtraction for chest radiography. *Appl Sci*. 2023;13(12):7220.

doi: 10.3390/app13127220

23. Rani G, Misra A, Dhaka VS, Zumpano E, Vocaturo E. Spatial feature and resolution maximization GAN for bone suppression in chest radiographs. *Comput Methods Programs Biomed*. 2022;224:107024.

doi: 10.1016/j.cmpb.2022.107024

24. Rani G, Misra A, Dhaka VS, *et al*. A multi-modal bone suppression, lung segmentation, and classification approach for accurate COVID-19 detection using chest radiographs. *Intell Syst Appl*. 2022;16:200148.

doi: 10.1016/j.iswa.2022.200148

25. Lam NFD, Sun H, Song L, *et al*. Development and validation of bone-suppressed deep learning classification of COVID-19 presentation in chest radiographs. *Quant Imaging Med Surg*. 2022;12(7):3917-3931.

doi: 10.21037/qims-21-791

26. Rajaraman S, Zamzmi G, Folio L, Alderson P, Antani S. Chest X-ray bone suppression for improving classification of tuberculosis-consistent findings. *Diagnostics* (*Basel*). 2021;11(5):840.

doi: 10.3390/diagnostics11050840

27. Rajaraman S, Cohen G, Spear L, Folio L, Antani S. DeBoNet: A deep bone suppression model ensemble to improve disease detection in chest radiographs. *PLoS One*. 2022;17(3):e0265691.

doi: 10.1371/journal.pone.0265691

28. Kim H, Lee KH, Han K, *et al*. Development and validation of a deep learning-based synthetic bone-suppressed model for pulmonary nodule detection in chest radiographs. *JAMA Netw Open*. 2023;6(1):e2253820.

doi: 10.1001/jamanetworkopen.2022.53820

29. Bae K, Oh DY, Yun ID, Jeon KN. Bone suppression on chest radiographs for pulmonary nodule detection: Comparison between a generative adversarial network and dual-energy subtraction. *Korean J Radiol*. 2022;23(1):139-149.

doi: 10.3348/kjr.2021.0146

30. Xu D, Xu Q, Nhieu K, Ruan D, Sheng K. An efficient and robust method for chest X-ray rib suppression that improves pulmonary abnormality diagnosis. *Diagnostics* (*Basel*). 2023;13(9):1652.

doi: 10.3390/diagnostics13091652

31. Thölke P, Mantilla-Ramos YJ, Abdelhedi H, *et al*. Class imbalance should not throw you off balance: Choosing the right classifiers and performance metrics for brain decoding with imbalanced data. *Neuroimage*. 2023;277:120253.

doi: 10.1016/j.neuroimage.2023.120253

32. Cohen JP, Dao L, Roth K, *et al*. Predicting COVID-19 pneumonia severity on chest X-ray with deep learning. *Cureus*. 2020;12(7):e9448.

doi: 10.7759/cureus.9448

33. Cohen JP, Morrison P, Dao L, Roth K, Duong TQ, Ghassemi M. COVID-19 image data collection: Prospective predictions are the future. *J Mach Learn Biomed Imaging*. 2020;2:1-38.

doi: 10.59275/j.melba.2020-48g7

34. Signoroni A, Savardi M, Benini S, *et al*. BS-Net: Learning COVID-19 pneumonia severity on a large chest X-ray dataset. *Med Image Anal*. 2021;71:102046.

doi: 10.1016/j.media.2021.102046

35. Borghesi A, Maroldi R. COVID-19 outbreak in Italy: Experimental chest X-ray scoring system for quantifying and monitoring disease progression. *Radiol Med*. 2020;125:509-513.

doi: 10.1007/s11547-020-01200-3

36. Borghesi A, Zigliani A, Masciullo R, *et al*. Radiographic severity index in COVID-19 pneumonia: Relationship to age and sex in 783 Italian patients. *Radiol Med*. 2020;125:461-464.

doi: 10.1007/s11547-020-01202-1

37. Caroprese L, Vocaturo E, Zumpano E. Argumentation approaches for explanaible AI in medical informatics. *Intell Syst Appl*. 2022;16:200109.

doi: 10.1016/j.iswa.2022.200109

38. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization. In *2017 IEEE International Conference on Computer Vision* (*ICCV*). Venice, Italy: IEEE; 2017:618-626.

doi: 10.1109/ICCV.2017.74

39. Musthafa MM, Mahesh TR, Vinoth Kumar V, Guluwadi S. Enhancing brain tumor detection in MRI images through explainable AI using Grad-CAM with resnet 50. *BMC Med Imaging*. 2024;24(1):107.

doi: 10.1186/s12880-024-01292-7

40. Talaat FM, Gamel SA, El-Balka RM, Shehata M, ZainEldin H. Grad-CAM enabled breast cancer classification with a 3D inception-ResNet V2: Empowering radiologists with explainable insights. *Cancers* (*Basel*). 2024;16(21):3668.

doi: 10.3390/cancers16213668

41. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional networks for biomedical image segmentation. In: *Medical Image Computing and Computer-Assisted Intervention-MICCAI 2015*. Cham: Springer; 2015:234-241.

42. *JSRT Database*. Available from: https://db.jsrt.or.jp/eng.php [Last accessed on 2024 Oct 15].

43. Isora P, Zhu JY, Zhou T, Efros AA. Image-to-image Translation with Conditional Adversarial Networks. In: *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*). Honolulu, HI, USA: IEEE; 2017:5967-5976.

doi: 10.1109/CVPR.2017.632

44. Zhu JY. Pytorch-CycleGAN-and-Pix2pix. Available from: https://github.com/junyanz/pytorch-cyclegan-and-pix2pix [Last accessed on 2024 Oct 15].

45. Horé A, Ziou D. Image Quality Metrics: PSNR vs. SSIM. In: *2010 20th International Conference on Pattern Recognition*. Istanbul, Turkey: IEEE; 2010:2366-2369.

doi: 10.1109/ICPR.2010.579

46. Wang Z, Bovik AC, Sheikh HR, Simoncelli EP. Image quality assessment: From error visibility to structural similarity. *IEEE Trans Image Process*. 2004;13(4):600-612.

doi: 10.1109/tip.2003.819861

47. Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely Connected Convolutional Networks. In: *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*). Honolulu, HI, USA: IEEE; 2017:2261-2269.

doi: 10.1109/CVPR.2017.243

48. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. In: *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*). Las Vegas, NV, USA: IEEE; 2016:770-778.

doi: 10.1109/CVPR.2016.90

49. Radosavovic I, Kosaraju RP, Girshick R, He K, Dollár P. Designing Network Design Spaces. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*); 2020:10428-10436.

doi: 10.48550/arXiv.2003.13678

50. Deng J, Dong W, Socher R, Li LJ, Li K, Fei-Fei L. ImageNet: A Large-scale Hierarchical Image Database. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*). Miami, FL, USA: IEEE; 2009:248-255.

doi: 10.1109/CVPR.2009.5206848

51. Schober P, Boer C, Schwarte LA. Correlation coefficients: Appropriate use and interpretation. *Anesth Analg*. 2018;126(5):1763-1768.

doi: 10.1213/ANE.0000000000002864

52. Karim KS, Tilley IS. Portable single-exposure dual-energy X-ray detector for improved point-of-care diagnostic imaging. *Mil Med*. 2023;188(Suppl 6):84-91.

doi: 10.1093/milmed/usad034

53. Shunkov YE, Kobylkin IS, Prokhorov AV, *et al*. Motion artefact reduction in dual-energy radiography. *Biomed Eng*. 2022;55:415-419.

doi: 10.1007/s10527-022-10148-9

54. Gulrajani I, Ahmed F, Arjovsky M, Dumoulin V, Courville A. Improved Training of Wasserstein GANs. In: *Neural Information Processing Systems 30* (*NIPS 2017*). San Diego, CA, USA; 2017:5769-5779.

55. Zhang L, Xu Z, Barnes C, *et al*. Perceptual Artifacts Localization for Image Synthesis Tasks. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision* (*ICCV*). Paris, France: IEEE; 2023:7545-7556.

doi: 10.1109/ICCV51070.2023.00697